

Clark JE, Osborne JW, Gallagher P, Watson S. [A simple method for optimising transformation of non-parametric data: an illustration by reference to cortisol assays](#). *Human Psychopharmacology: Clinical and Experimental* 2016, **31**(4), 259-267.

**Copyright:**

This is the peer reviewed version of the following article: Clark JE, Osborne JW, Gallagher P, Watson S. [A simple method for optimising transformation of non-parametric data: an illustration by reference to cortisol assays](#). *Human Psychopharmacology: Clinical and Experimental* 2016, **31**(4), 259-267., which has been published in final form at <http://dx.doi.org/10.1002/hup.2528> This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.

**Date deposited:**

06/09/2016

**Embargo release date:**

27 May 2017



This work is licensed under a [Creative Commons Attribution-NonCommercial 3.0 Unported License](#)

# A simple method for optimising transformation of non-parametric data: an illustration by reference to cortisol assays.

---

James E Clark<sup>1</sup>, MA(Hons), MBPsS

Jason W Osborne<sup>2</sup>, BA, MA, PhD, PStat(r)


Peter Gallagher<sup>1</sup>, Bsc(Hons), MPhil, PhD, AFBPsS

Stuart Watson<sup>1</sup>, MBBS, MD, MRCPsych 

<sup>1</sup>The Institute of Neuroscience, Newcastle University, UK

<sup>2</sup>The Graduate School and Department of Mathematical Sciences, Clemson University, Clemson, SC, USA

## Corresponding Author:

 Stuart Watson, Academic Psychiatry and Regional Affective Disorders Service, The Institute of Neuroscience, Wolfson Research Centre, Campus of Ageing and Vitality, Newcastle University, Newcastle Upon Tyne, NE4 5PL, UK. Tel.: +44 191 208 1388. Fax: +44 (0) 191 256 3324. E-mail Address: [stuart.watson@ncl.ac.uk](mailto:stuart.watson@ncl.ac.uk)

## 1. Abstract

Neuroendocrine data is typically positively skewed and rarely conforms to the expectations of a Gaussian distribution. This can be a problem when attempting to analyse results within the framework of the general linear model, which relies on assumptions that residuals in the data are normally distributed. One frequently used method for handling violations of this assumption is to transform variables to bring residuals into closer alignment with assumptions (as residuals are not directly manipulated). This is often attempted through ad hoc traditional transformations such as square root, log, and inverse. However, Box and Cox (1) observed these are all special cases of power transformations, and proposed a more flexible method of transformation for researchers to optimize alignment with assumptions. The goal of this paper is to demonstrate the benefits of the infinitely flexible Box-Cox transformation on neuroendocrine data using syntax in SPSS. When applied to positively skewed data typical of neuroendocrine data, the majority (~2/3) of cases were brought into strict alignment with Gaussian distribution (i.e., a non-significant Shapiro-Wilks test). Those unable to meet this challenge showed substantial improvement in distributional properties. The biggest challenge were distributions with a high ratio of kurtosis to skewness. We discuss how these cases might be handled, and we highlight some of the broader issues associated with transformation.

## 2. Introduction

The general linear model (GLM) is of fundamental importance in neuroendocrine research, and forms the foundations of basic analytic tools such as t-tests, analysis of variance (ANOVA) and linear regression. However, its application relies on the assumption that the residuals in the model conform to the Gaussian (normal) distribution.

The shape of the distribution can be characterised according to its standardised moments, which are the ratio of the central moments to the standard deviation (2). The skewness and kurtosis of a data set are the 3<sup>rd</sup> and 4<sup>th</sup> standardised moments respectively (3, 4). A useful heuristic for visualising skew may be to regard it as the standardised difference of the mean from the mode (5) (6), hence,

the normal distribution has skewness of 0. Kurtosis is considered to be a vague concept to visualise (7) and is often poorly described in textbooks. It is perhaps helpful to visualise in terms of the dispersion of the  $Z^2$  distribution about 1 (8) and is a measure of the shape of the distribution tails; high kurtosis represents strong central tendency and lower values represent heavy tails (9).

Accordingly, some commentators also frame kurtosis in terms of the “shoulder” size of a distribution (10). Student (11) used the images in Figure 1 as a novel way of visualising kurtosis. The lower bound of kurtosis is 1 and occurs in a bimodal distribution though there is no upper bound. The normal distribution has kurtosis of 3, prompting researchers to classify *excess kurtosis* (defined simply as kurtosis minus 3 in SPSS) which is 0 in the normal distribution and the minimum value becomes -2. This is an important distinction to bear in mind to avoid potential confusion, particularly as these two measures are often used interchangeably in analytic software. The Shapiro-Wilk statistic (W), essentially computes the ratio of the data parameters under normal assumptions to those in the real data and tests the null hypothesis that this ratio is 1 (12). It offers a convenient formal test of normality which is easily available in most statistical packages. The Shapiro-Wilk statistic is considered throughout this text, but the reader is reminded that there are a large number of other distribution statistics, including Kolmogorov-Smirnov, Anderson-Darling, and the Pearson Goodness-of-Fit Test. Also note particularly that these tests tend to be exceptionally sensitive to minimal (i.e., unimportant or ignorable) deviations from the normal distribution, particularly when power (sample size) gets larger.

When a particular data set does not meet the assumptions required for GLM analysis, ranking tests may be used; though such tests lose value when dealing with large samples or when multiple ranking ties are present (13). Furthermore, they do not offer the same flexibility in experimental design that the GLM offers (14). Another strategy is to remove outliers or inappropriately influential cases from a data set (e.g. cases with relatively large residuals, leverage, or influence). This approach however may be seen as subjective (15) and potentially engenders a reduction in power in very small sample sizes (16). This is particularly pertinent in clinical neuroendocrine research where large

samples can be difficult to obtain. It may, however, be useful if one highly extreme value is severely altering the distribution in question and more conservative data trimming methods, such as Winsorization, may be of benefit (17). However, it is generally more useful to transform non-parametric neuroendocrine data in order to bring skewness and excess kurtosis as close as possible to zero so that the data (or more often, the residuals) is more normally distributed. To be clear, it is not possible at this time to directly manipulate the distribution of residuals when assumptions are violated, and thus, transformation of the observed variables is a valuable yet indirect method to improve compliance with assumptions.

Transforming a data set can also permit analysis within different types of model. For example, an additive model assumes no interaction term and shows that the effect of predictor variables remain constant across changes in others. In contrast multiplicative models contain interaction effects which are modelled as the product of two or more of the predictors. Logarithmic transformations are very useful for inducing additivity into multiplicative models, though clearly the opposite will also be true; that transformation by raising a sample to a given power can induce interaction effects in previously additive models.

Transformations are normally computed manually through a trial and error process, in which a small collection of well-known functions are applied to the data. Such functions are  $x^2$ ,  $\sqrt{x}$ ,  $1/x$ ,  $\log x$  or  $\ln x$ . The transformed data is then analysed using W. If these transformations fail to normalise the data the researcher must then decide whether to use ranking tests or simply use the GLM regardless. Of course it may also be the case that even if the data is normalised via one of these arbitrarily chosen transformations it is possible that some other transformation would yield a closer approximation to the normal distribution. It should be noted that all transformations other than logarithmic ones can be represented in the general form  $x^\lambda$ . The Box-Cox transformation offers a more formal and highly flexible method of transformation to optimally improve the likelihood of the data given specific distribution model (e.g. Gaussian) (1). The “MASS” and “AID” packages in R can be utilised to apply the Box-Cox transformation, however, this transformation remains largely unused in

neuroendocrine research, perhaps due to its computational complexity and its absence from many analytic packages. As such, Osborne (18) described a rapid and simple method for determining the optimum transformation for any particular non-parametric data set *a posteriori* using syntax in SPSS, arguably the Statistical Package which is most widely used for clinical neuroendocrine data. We present here a more detailed discussion of this procedure alongside examples of its application. We then examine situations in which the procedure is unsuccessful. This will allow the formulation of general rules of thumb with which to approach analysis and aid researchers in determining the outcome of transformations.

### 3. The Transformation Procedure

This procedure is set out in Appendix 1 and is based on that of Osborne(17, 18) which was designed for determining optimal parameters for the Box-Cox transformation though here we present a more detailed description of its foundations alongside suggestions for initial parameters and cut off scores which are useful in predicting its success. We will consider a 1xN column vector,  $X$ , which relates to N participant's scores on some variable  $x$ . This is in order to conform to the way in which data is presented in SPSS. We will describe a method for determining the optimum value of  $\lambda$  in the Tukey Transformation:

$$f(x) = x^\lambda \quad (19)$$

The first step in the transformation procedure is to anchor the minimum value of  $X$  to 1 via the following:

$$X + (1 - X_{min})$$

where  $X_{min}$  is a 1xN column vector containing only  $x_{min}$ . Osborne (20) has shown in a very detailed way that this minimises residual skew in the transformed data, with the obvious benefit deriving from  $1^\lambda = 1$  for all values of  $\lambda$ .

It is then necessary to specify the number of different values of  $\lambda$  ( $K$ ) we wish to test using a VECTOR command which creates discrete variables in the SPSS data viewer corresponding to  $x^{\lambda_1, \dots, \lambda_K}$ . The

LOOP command is used to ensure the transformation will be performed from  $\lambda_{1...k}$  with a constant ( $\alpha$ ) and an interval by which the value of  $\lambda$  increases after each loop of the function ( $\beta$ ). As such, the value of  $\lambda$  after the  $i^{\text{th}}$  loop is determined by the following:

$$\lambda = \alpha + i\beta$$

It is important to note that  $\lambda$  is undefined for  $i = 0$  and so  $\lambda_1$  will always take the form:

$$x^{(\alpha+\beta)}$$

We also specify a transformation for the arbitrary case of  $\lambda = 0$  using the DO IF command. In keeping with Box and Cox (1) the natural logarithm (ln) is used. Logarithmic-transformations are chosen because they are particularly useful for transforming positive skew and because of their ability to linearize models that have a multiplicative functional form. Thus the statistical comparison of two groups is based on an easily interpretable effect measure, i.e. the ratio of the geometric means between the two contrasted groups (scaled in relative instead of absolute change), though other functions can be specified if necessary.

All that is left is to specify the transformation which in this case, as stated previously, is  $x^\lambda$ , and enter the appropriate values into the “frequencies” and “examine” commands. SPSS will then compute K new variables corresponding to the pre-defined values of  $\lambda$ , and offer descriptive statistics, histograms, Q-Q and box plots and W for each transformed variable. The complete syntax is presented in the appendix 1.

It is also useful to set in place general rules of thumb with which to approach the transformation procedure in order to determine an appropriate range of values of  $\lambda$  to be tested. This will reduce the number of computations required and maximise simplicity in the SPSS data viewer and output. In formulating these rules of thumb it is necessary to remember that the best transformation is the one that brings skew and excess kurtosis as close as possible to zero. Different ranges of  $\lambda$  will therefore be more appropriate for differently skewed data and data with different kurtoses.

In the case of positively skewed data, values of  $\lambda$  greater than 1 or less than -1 will be of no use. We wish to reduce the difference between the mean and the mode, but this will exponentially increase

at values greater than 1 and will therefore only increase skew. Values below -1 are likely to exponentially increase lower values of  $X$  and simply increase skew again. This is illustrated graphically in Figure 2a. With the analogous definition of skewness and kurtosis in mind (according to the standardised moments of the distribution) the same values of  $\lambda$  would be expected to be useful in transforming high kurtosis, as shown in Figure 2b. Our experience suggest that non-parametric neuroendocrine data is frequently positively skewed(21-23), indeed of the 91 data sets here examined, 90 were positively skewed, though kurtosis varies without a discernible pattern. In order to reduce the skewness of data we recommend the parameters:  $\alpha = -0.05$ ,  $K = 21$ ,  $\beta = 0.05$ . In the case of negative skew,  $\lambda$  should be greater than 1. This is displayed in Figure 2c, which is a graph from the only non-normal case of negative skew which was available. As such, we recommend that the parameters of the transformation procedure are set to:  $\alpha = 0.5$ ,  $K = 11$ ,  $\beta = 0.5^2$ . It is important to note that, although it is discussed here, this is an extremely rare finding amongst neuroendocrine data.

#### **4. Application and exploration of the method**

In order to demonstrate this general strategy in transforming a single data set, we will compare the characteristics of several positively skewed samples and so establish more precisely when particular transformations may be more useful.

The sample used in this example is a measure of plasma cortisol taken at 13:30 after orally administered dexamethasone at 13:00 in a sample of 18 participants with bipolar affective disorder. It was taken as part of a placebo controlled investigation into the effects of fludrocortisone and dexamethasone on endocrine and neuropsychological function. The raw distribution of the data is shown in Figure 3a. The skewness was 1.29 and excess kurtosis was 2.47. Correspondingly, the Shapiro-Wilk statistic ( $W$ ) = .916,  $P < .01$ . The minimum value of the sample was 54 and subsequently 53 was subtracted from every data point in the distribution in order to anchor the minimum value in the data set to 1. The parameters were set at  $\alpha = -0.05$ ,  $K = 21$ ,  $\beta = 0.05$ . The transformation revealed



that when  $\lambda = 0.55$  the data was closest to normal (Figure 3b). The skewness after transformation was -0.1 and excess kurtosis was 1.4. As such  $W = 0.98$ ,  $P = 0.68$ . This transformation has therefore greatly reduced the skewness and kurtosis and consequently the distribution is no longer significantly different from normal. (Note that Shapiro-Wilk, and other similar hypothesis tests of distributional assumptions, are highly sensitive to small deviations from the distributional ideal, particularly when power increases through relatively larger sample size. Thus, ignorable deviations from strict Gaussian distributions might yield a rejected null hypothesis when performed in the context of even moderately large samples, as you will see below. Caution must be exercised when interpreting the results of an S-W test; examining improvement in distributional parameters is also valuable).

In the 91 endocrine data sets that we examined (see legend of figure 2 for details), the range of skewness was much lower (0.6-8.4) than that of excess kurtosis (-0.5-70.8); this is a typical finding in neuroendocrine data. We ran the transformation procedure according to the parameters suggested above and found that in 60 samples (~66%) the transformed distribution no longer resulted in significant deviations from normality, as measured by  $W$ . The characteristics of the transformed and non-transformed samples are shown in Table 1. Note that in most cases, even when the S-W test did not result in a retained null hypothesis, distributional qualities improved, and thus should not necessarily be considered a “failure” of the transformation, but rather a limitation of the S-W test itself. In order to examine possible reasons behind the failure of the transformation procedure in the cases which were unsuccessful, we tested the hypothesis that failure resulted from two distinct possible scenarios:

- If the size of a sample was very large
- If kurtosis and skewness were both high and kurtosis was proportionally larger than skewness

The second of these follows from the assumption that kurtosis and skewness are both transformed using the same parameters and thus if the kurtosis of a sample is proportionally larger or smaller

than its skewness any data transformation may be less likely to succeed. We therefore created the computed variable excess kurtosis/skewness (K/S). In order to test these hypotheses, we conducted a binary logistic regression to examine the effects of sample size and K/S on the likelihood of a data set being subsequently transformed. The outcome variable was the log odds ratio of a sample being transformed vs not-transformed according to the Shapiro Wilk test. The model was significant;  $\chi^2=27.8$ ,  $P<0.001$  and explained 36% of variance in outcome according to the Nagelkerke  $R^2$  statistic (24). In total the model correctly classified 79% of cases as “transformed” or “not-transformed” and a higher K/S ( $P<0.05$ ) and higher sample size ( $P<0.001$ ) were both associated with a greater likelihood of the transformation procedure failing, as measured by W. Wald chi-square showed that the effect of sample size was larger ( $\chi^2=13.5$ ) than that of K/S ( $\chi^2=4.8$ ) in determining group membership. Receiver operating characteristics showed that a K/S higher than 2.7 showed 82% sensitivity and 77% specificity in classifying samples as “non-transformed” according to the regression model, whilst a sample size of 49 had 94% sensitivity and 92% specificity of being in the same category. We propose that, as a general rule of thumb, failure of the transformation may result when a sample meets these criteria (i.e.  $k/s>2.7$ , sample size of  $\geq 49$ ).

Of course, until now we have only considered a case in which we wish to transform an individual sample; but the majority of transformations aim to facilitate analysis of multiple samples within ANOVA or linear regression. This is simply a logical extension of the process described above. If this procedure is performed on all the samples which are to be analysed the optimum value of  $\lambda$  can be determined using the maximum value of W at which normal assumptions are met in each sample. It is possible to transform distributions over each variable separately, but this makes comparison slightly trickier due to their different measurement scales. As a straightforward heuristic it is simpler to use an identical transformation across all variables in a repeated measures design. We will illustrate how to find the optimum transformation of multiple endocrine samples using data gathered from the cortisol awakening response (25); collected at 5 points from wakening until 1 hour post-wakening. Table 3 shows the results of transformation after concurrent values of  $\lambda$  as well

as the mean  $W$  for each value. It can be seen that a value of  $\lambda = .00$  normalised each sample and provided the greatest mean value of  $W$ . This is, therefore, the most useful value with which to transform this particular family of samples.

Finally, it is necessary to describe the procedure in the case of negative skew, although in this field such a distribution may be rare (supporting this perspective, only one of our samples showed negative skew)finding. Regardless, the power of the Box-Cox transform is that it can capably handle both positive and negatively skewed variables by adjusting the parameter. The AUC measure with respect to increase (AUCi) (26), in which cortisol samples had been taken from midday until night, . It is displayed graphically in Figure 3c. As can be seen there was a general decrease in cortisol, though the substantially greater decrease of a few participants rendered the distribution negatively skewed. Whilst we have acknowledged the rarity of this example, such distributions may be garnered from AUCi measures and it is important to show how the transformation procedure can help handle this. This sample had a skewness of -1.5 and excess kurtosis of 2.1.  $W = 0.851$ ,  $P < 0.001$ . Setting the procedure according to the parameters recommended above, it was shown that the optimum transformation occurred at  $\lambda = 2.5$  (Figure 3d). Skewness was now 0.04 whilst excess kurtosis was -0.39.  $W = 0.961$ ,  $P = 0.28$ . Given the lack of available data, it is not possible to offer an examination of the characteristics of samples which can be transformed using this procedure, though this provides a useful example of strategies which can help transform negative skew. Note that the SPSS macro from Osborne (20) is sufficiently flexible that you can simply specify a range of 60 or more  $\lambda$  (e.g., -3.0 to 3.0 in increments of 0.10) which can easily handle either positive or negatively skewed variables. Given modern desktop computing power, selection of  $\lambda$  is not an issue a researcher should worry over.

## 5. Discussion

In this paper we have described a transformation procedure in SPSS syntax for transforming non-parametric data in order to facilitate analysis within the GLM. We then gave detailed examples of its

applications and showed that it was successful in transforming around two thirds of the non-parametric data we had available. Note that in this case, we defined “successful” as a relatively high bar of a resulting distribution that was not significantly different from normal. However, normality is a continuous variable (27), and failure to reach that threshold, particularly in relatively large samples, does not necessarily mean the transform has failed. Material improvement in the distribution, and in particular, in meeting the assumptions of the statistical procedure, is valuable even absent a non-significant statistical test of normality.

Not surprisingly, the most common reason for a transform to fail to meet our criteria was sample size. It is important to note that in the case of relatively large sample size, failure of the transformation procedure is likely due to the limitations in using  $W$ , which is very highly powered to detect even small deviations from normality at sample sizes above 50 (28). Incidentally this is very close to our derived cut off value for sample size. If sample size is large, therefore, even transformations which greatly reduce skew may be rendered non-normal using  $W$ . For example, one of our samples had an initial skewness of 2.4 and excess kurtosis of 6.3.  $W = .663$ ,  $P < .001$  (Figure 3e). Applying the transformation procedure, we found that log transformation reduced skew to .29 and excess kurtosis to -.51.  $W = .962$  (Figure 3f). Clearly the procedure has worked very well, although the sample size has powered the Shapiro-Wilk test to such an extent that the distribution was still classed as non-normal. In this case it may be more appropriate to consider Q-Q Plots or simply examine the histogram of the distribution by eye, as it is likely to be adequate for analysis within the GLM (29).

Given a reasonably continuous variable (discontinuous, bimodal, and discrete variables are not appropriate for transformation of this nature), it is unlikely that this procedure will fail to improve distributional quality.

This allowed us to formulate general rules of thumb to refine the procedure before it is implemented and suggest parameters which must be met in order for the procedure to be successful. This method was then extended to give an example of transforming several samples for

comparative analysis. In the case of negative skew we gave an estimate for parameters in the SPSS syntax and showed how these can successfully transform a particular sample, though we could not obtain enough negatively skewed samples to explore situations in which these parameters may be inadequate.

Miller and Plessow (30) have recently provided a comprehensive overview on the use of transformation methods in neuroendocrine research which they conclude is a necessary prerequisite for analysis within the GLM. They show that different power transformations are often appropriate with different types of experimental data and offer simply rules of thumb which can be applied in different scenarios. Our results build upon this work and offer a more bespoke method for determining the optimum transformation for any given data set, using a simple syntax in SPSS. We also discuss instances in which transformation may prove inappropriate which may provide a useful guide in initial approaches to data analysis.

The method that we describe is a minor development of Osborne's translation of the Box-Cox transformation into SPSS. It allows the identification of the optimum transformation factor. One potential disadvantage of this data specific strategy is that the range of transformations used in neuroendocrine literature will increase which may engender challenges in comparing findings between studies and in conducting meta-analyses. This is particularly relevant given the relative robustness of ANOVA to slight deviations from normality (29).

It is also important to discuss whether transformation of data is a legitimate approach for analysis. From a purely statistical standpoint, transformation is perfectly acceptable but a key point is whether the "outliers" in a data set represent a genuine property of the population being measured or whether they represent sampling error. For example, the literature on reaction time in attentional performance generally shows responses are heavily skewed, though the tail of the distribution contains information regarding slower, but more controlled and strategic processes. In this case, modelling the data in an ex-Gaussian distribution (which is the convolution of the exponential and Gaussian distributions) has proved extremely beneficial (31). Similarly, incorporating skewness into

distribution parameters has proven useful in the general medical literature (32). In relation to this question, Miller & Plessow suggest that non-linearity in neuroendocrine data is the result of the specific nature of oscillations within such systems (30). The question is then posed as to whether the skewness which is frequently seen in neuroendocrine data is an abnormality which should be normalised or whether it is an integral part of the endocrine response which contains valid information and should be modelled accordingly.

One other potential limitation regarding ANOVA comparison which is raised in the literature is that transformation may remove linearity and additivity of the resultant model (33). In turn, this can lead to erroneous conclusions regarding interaction effects in factorial analysis (34). It should be noted that log transformations can be useful in removing non-linearity from an analysis if this is desired. Tukey's test of additivity (35) tests the null hypothesis that the interaction coefficient in the ANOVA model is zero and may prove useful in assessing whether changes in linearity have occurred as a result transformation. Specifically, additivity should be checked before and after transformation and results interpreted accordingly if any changes have occurred.

It is also important to note that other methods for handling non-parametric data are available and, as in transformation methods, each is associated with certain benefits and costs. For example, bootstrapping small sample sizes may reduce bias in parameter estimation (36) and generalized linear models have been developed which allow for modelling of various different outcome distributions within a linear framework (37). Similarly, Bayesian inferential methods allow for the use of a wide variety of distributions in outcome variables based on estimates in prior distributions (38). Nevertheless, transformation methods still remain a useful alternative and are more easily implemented and practically straightforward, making them a useful alternative for researchers in neuroendocrinology.

## 6. Conclusions

Neuroendocrine data frequently is positively skewed with high kurtosis and as such fails to meet the assumptions required for ANOVA. It is our opinion that such data sets may not always be optimally transformed. We recommend the use of Osborne's (18) transformation procedure and its associated SPSS syntax as an effective, but practically simple method for achieving this. We present a variation of Osborne's (18) syntax and seek to highlight the ready availability of considered transformation strategies for neuroendocrine data.

## Appendix 1

**The SPSS Syntax** (adapted from Osborne, 2010)

```
COMPUTE var1=  $x + (1 - x_{min})$ .
```

```
EXECUTE.
```

```
VECTOR lam1(K)/x1(K).
```

```
LOOP idx=1 TO K.
```

```
- COMPUTE lam1 (idx)= $\alpha + \text{idx} * \beta$ .
```

```
- DO IF lam1 (idx)=0.
```

```
-   COMPUTE x1 (idx) =LN(var1).
```

```
- ELSE.
```

```
-   COMPUTE x1 (idx)= (var1**lam1(idx)).
```

```
- END IF.
```

```
END LOOP.
```

```
EXECUTE.
```

```
FREQUENCIES VARIABLES var1 x11 to x1K
```

```
  /FORMAT NOTABLE
```

```
  /STATISTICS SKEWNESS KURTOSIS
```

```
  /ORDER ANALYSIS.
```

```
EXAMINE VARIABLES var1 x11 to x1K
```

```
  /PLOT BOXPLOT HISTOGRAM NPLOT
```

```
  /COMPARE GROUPS
```

```
  /STATISTICS DESCRIPTIVES
```

```
  /CINTERVAL 95
```

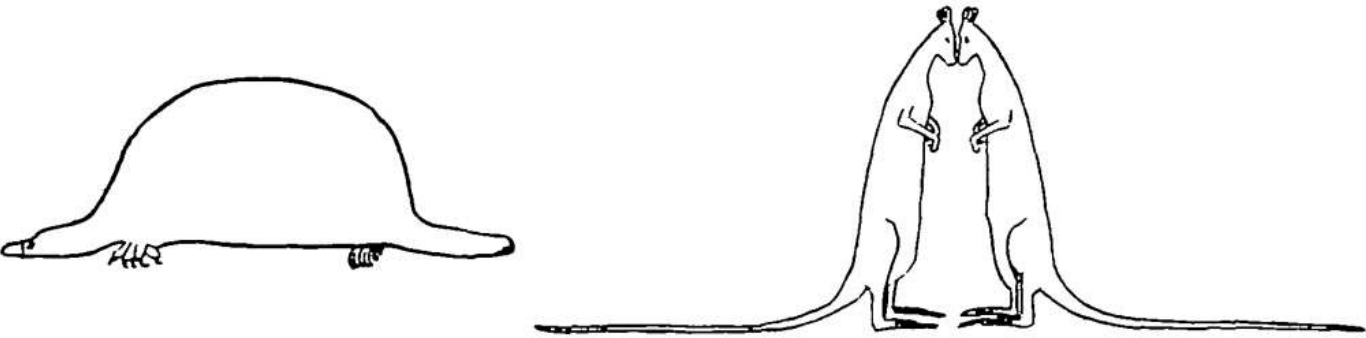
```
  /MISSING LISTWISE
```

```
  /NOTOTAL.
```

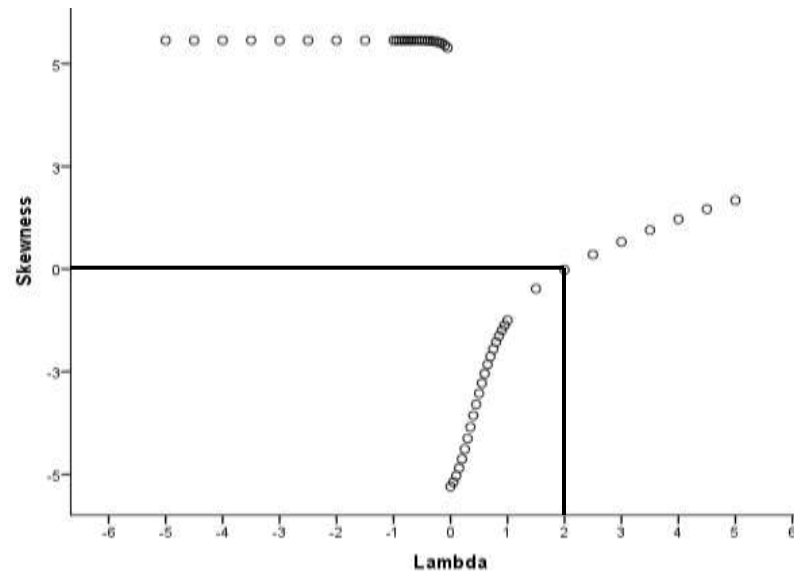
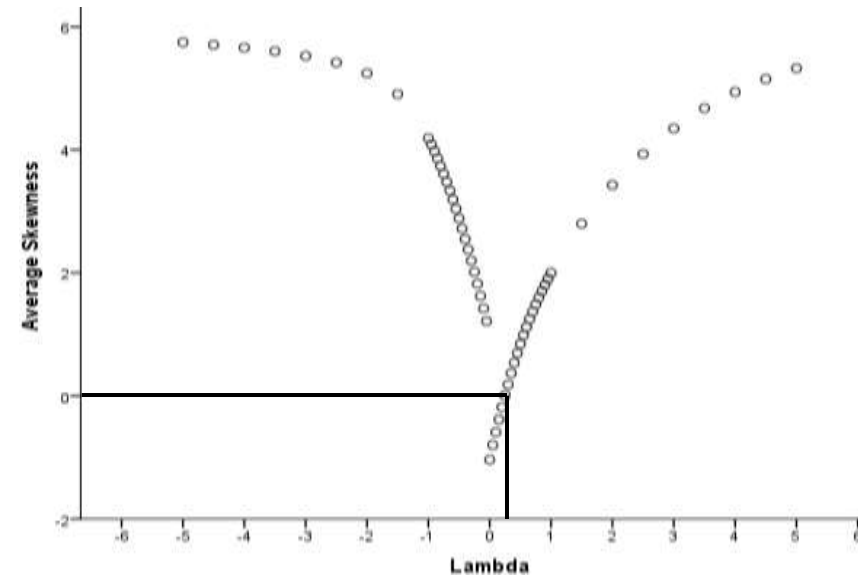
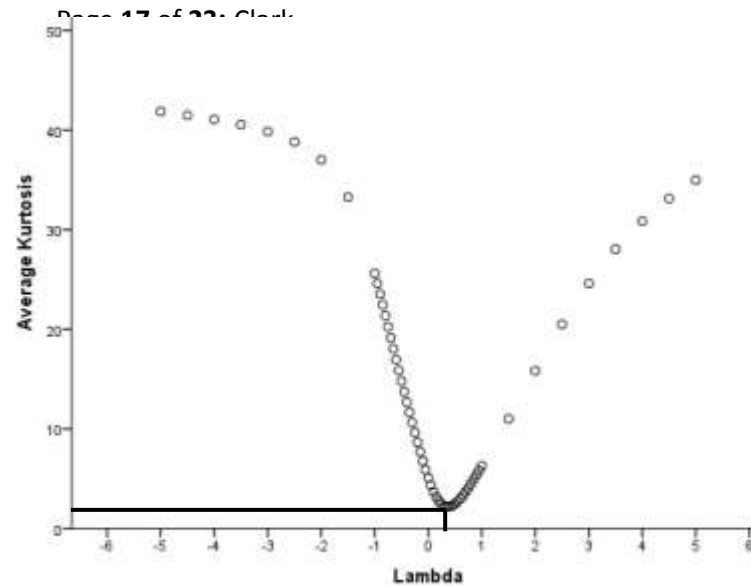


## List of Figures and Tables

**Figure 1-** “Student's” demonstrations of low and high kurtosis



he kangaroos on the right



### Legend Figure 2

2a and 2b show how the average kurtosis and skewness of the 91 data sets varied with different values of lambda.

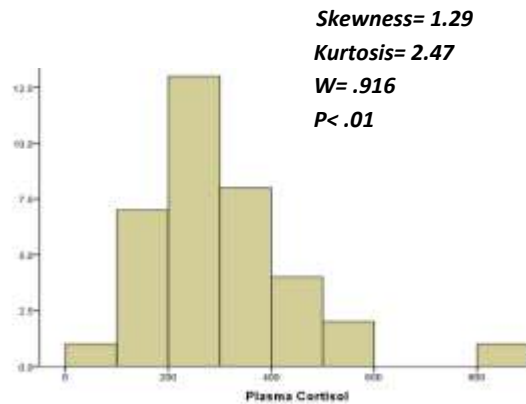
The data was taken from three studies. The first was an investigation into the endocrine and neuropsychological effects of a mineralocorticoid receptor agonist and a glucocorticoid receptor agonist in euthymic patients with bipolar disorder and healthy controls. The second was an examination of the classification accuracy of the dexamethasone suppression test and the dex/crh test in a variety of mood disorders. The third was a randomised control trial of the drug mifepristone in depressed bipolar patients. In total 91 samples were available which were positively skewed. These comprised a variety of measures including saliva and plasma responses to pharmacological challenge, awakening responses and area under the curve (AUC) measures.

The average kurtosis was positive, despite its large variability, as was skewness.

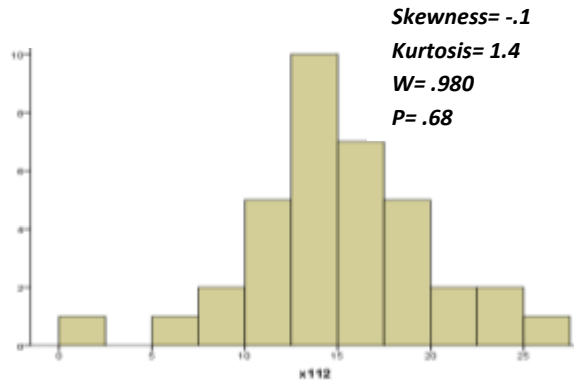
Principle findings in both cases were that the optimal value of lambda was between 0 and 1. Of course, these graphs are averages and so there will be cases when these rules of thumb are not appropriate but they are the most likely to work on positively skewed data.

2c is a graph showing the change in a negatively skewed sample with various values of lambda. The sample is an AUC measure of plasma cortisol derived from samples taken over a period of 7 hours from midday until night. The graph shows that a value greater than 1 is best for optimising skew.

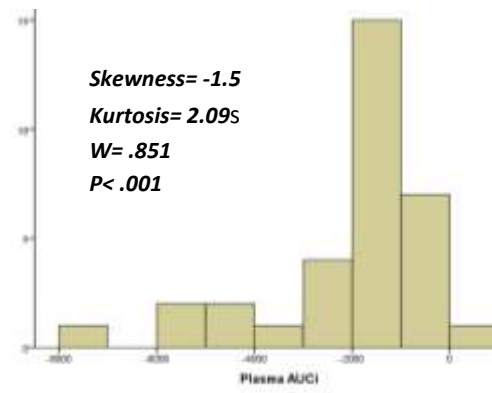
**Figure 3a-** Histogram of plasma cortisol showing the positively skewed nature of the distribution



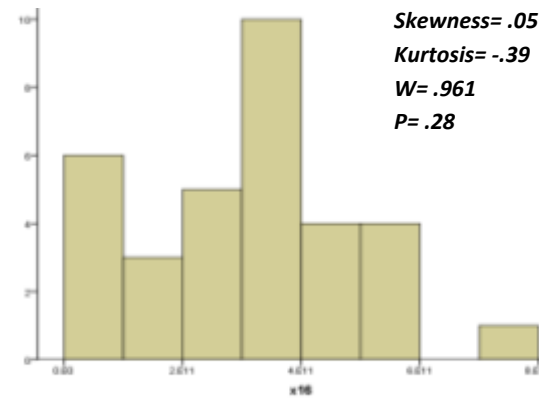
**Figure 3b-** Histogram of the cortisol data after the 12<sup>th</sup> iteration (x112) of the transformation procedure



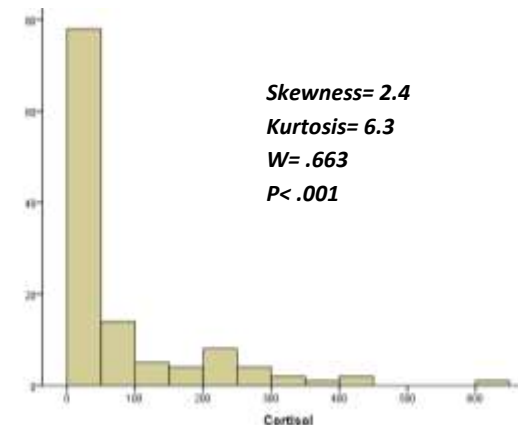
**Figure 3c-** Histogram showing negatively skewed AUC<sub>i</sub> measure



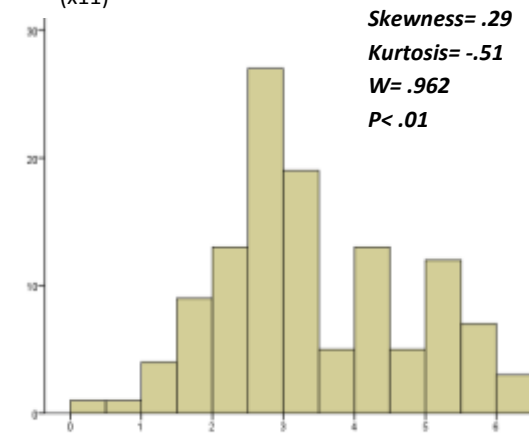
**Figure 3d-** Histogram showing the distribution of the transformed data after the 6<sup>th</sup> iteration (x16)



**Figure 3e-** Histogram showing positively skewed distribution, with a large sample size



**Figure 3f-** Histogram showing the distribution of the transformed data after the 1<sup>st</sup> iteration (x11)



### Legend Figure 3

3a shows the distribution of sample of plasma cortisol taken at 13:30 after orally administered dexamethasone at 13:00 in a sample of 18 participants with bipolar affective disorder. It was taken as part of a placebo controlled investigation into the effects of fludrocortisone and dexamethasone on endocrine and neuropsychological function. The distribution has noticeable positive skew and kurtosis and is significantly different from normal according to W.

3b shows the distribution of the data after transformation with the optimum value of lambda- 0.55.

3c shows the distribution of a measure of AUC<sub>i</sub> taken from sequential measures of plasma cortisol. There is clear negative skew and high kurtosis.

3d shows the transformed distribution after the 6<sup>th</sup> iteration of the transformation procedure, which was successful in normalising the raw data.

3e shows the distribution of a measure of plasma cortisol which contained 118 individuals. Skewness is strongly positive and kurtosis is very high.

3f shows the distribution after log transformation. The skewness and excess kurtosis have both been brought very close to zero, though W still suggests the distribution is not normal.

**Table 1- Characteristics of subsequently transformed and non-transformed cortisol samples. Descriptive and inferential statistics are displayed.**

	Transformed		Not-Transformed		Comparison <sup>1</sup>
	Mean	SD	Mean	SD	d
Skewness	1.74	1.13	2.68	1.58	.68
Kurtosis	4.96	7.64	10.75	13.84	.52
K/S	1.96	1.58	3.10	1.75	.68
W	.83	.13	.67	.18	1.43
N	39.23	15.00	70.35	39.58	1.04

W= *Shapiro-Wilk* statistic, N= Sample size, K/S= Kurtosis/Skewness, d= Cohen's d (0.2= small effect, 0.5= medium effect, 0.8= large effect)

<sup>1</sup> Comparison of samples which could with those which couldn't be transformed

#### Legend Table 1

Principal findings are:

- . Samples which were transformed by the procedure tended to lower skewness and kurtosis
- . Further to this, samples which were transformed also had a lower ratio of kurtosis to skewness
- . The Shapiro-Wilk statistic was also much higher in the transformed data
- . In general the sample size of transformed data was also much lower

**Table 2- Table showing the performance of each transformation across five different measures to be analysed.**

$\lambda$	W after each transformation											
	1		2		3		4		5		Mean	
	W	P	W	P	W	P	W	P	W	P	W	P
0.00	.968	.376	.964	.313	.950	.107	.962	.243	.983	.830	.965	.374
0.05	.972	.478	.960	.239	.945	.074	.957	.170	.981	.778	.963	.348
0.10	.975	.570	.952	.141	.938	.043	.949	.095	.979	.701	.959	.310
0.15	.977	.642	.941	.064	.928	.022	.938	.043	.976	.600	.950	.274
0.20	.978	.690	.925	.023	.916	.010	.923	.016	.972	.484	.943	.245
0.25	.979	.712	.906	.007	.901	.004	.906	.005	.968	.366	.932	.219
0.30	.979	.709	.884	.002	.885	.001	.886	.001	.963	.259	.919	.194
0.35	.978	.682	.858	.000	.866	.000	.863	.000	.957	.172	.904	.171
0.40	.977	.631	.830	.000	.845	.000	.838	.000	.951	.108	.888	.148
0.45	.974	.559	.799	.000	.822	.000	.810	.000	.943	.065	.870	.125
0.50	.972	.472	.767	.000	.798	.000	.781	.000	.936	.037	.851	.102
0.55	.968	.379	.733	.000	.773	.000	.751	.000	.927	.021	.830	.080
0.60	.964	.289	.699	.000	.747	.000	.720	.000	.918	.011	.810	.060
0.65	.960	.210	.665	.000	.720	.000	.689	.000	.909	.006	.789	.043
0.70	.955	.146	.631	.000	.693	.000	.657	.000	.899	.003	.767	.030
0.75	.949	.097	.598	.000	.666	.000	.626	.000	.888	.002	.745	.020
0.80	.943	.063	.565	.000	.638	.000	.595	.000	.877	.001	.724	.013
0.85	.937	.040	.535	.000	.611	.000	.566	.000	.865	.000	.703	.008
0.90	.930	.025	.505	.000	.585	.000	.537	.000	.853	.000	.682	.005
0.95	.923	.015	.478	.000	.559	.000	.510	.000	.841	.000	.662	.003
1.00	.915	.009	.452	.000	.535	.000	.484	.000	.828	.000	.643	.002

#### Legend Table 2

The mean value of W is the mean of the 5 samples after transformation with each value of lambda. Transformations which normalised all measures according to W are highlighted. The maximum value of W when all samples are normalised occurs when lambda is 0.00.

## References

1. Box GEP, Cox DR. An analysis of transformations. *Journal of the Royal Statistical Society, Series B.* 1964; 26(2): 211-52.
2. Oja H. On location, scale, skewness and kurtosis of univariate distributions. *Scandinavian Journal of Statistics.* 1981: 154-68.
3. Groeneveld RA, Meeden G. Measuring skewness and kurtosis. *The Statistician.* 1984: 391-9.
4. DeCarlo LT. On the meaning and use of kurtosis. *Psychological Methods.* 1997; 2(3): 292.
5. Pearson K. Contributions to the mathematical theory of evolution. II. Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London A.* 1895: 343-414.
6. Arnold BC, Groeneveld RA. Measuring skewness with respect to the mode. *The American Statistician.* 1995; 49(1): 34-8.
7. Mosteller F, Tukey JW. *Data analysis and regression: a second course in statistics.* Addison-Wesley Series in Behavioral Science: Quantitative Methods. 1977.
8. Moors JJA. The meaning of kurtosis: Darlington reexamined. *The American Statistician.* 1986; 40(4): 283-4.
9. Balanda KP, MacGillivray HL. Kurtosis: A Critical Review. *The American Statistician.* 1988; 42(2): 111-9.
10. Liang Z, Wei J, Zhao J, Liu H, Li B, Shen J, et al. The statistical meaning of kurtosis and its new application to identification of persons based on seismic signals. *Sensors.* 2008; 8(8): 5106-19.
11. "Student". *ERRORS OF ROUTINE ANALYSIS.* *Biometrika.* 1927; 19(1-2): 151-64.
12. Shapiro SS, Wilk MB. An analysis of variance test for normality(complete samples). *JSTOR,* 1964.
13. Whitley E, Ball J. Statistics review 6: Nonparametric methods. *CRITICAL CARE-LONDON.* 2002; 6(6): 509-13.
14. Gaito J. Non-parametric methods in psychological research. *Psychological Reports.* 1959; 5(g): 115-25.
15. Orr JM, Sackett PR, Dubois CL. Outlier detection and treatment in I/O psychology: A survey of researcher beliefs and an empirical illustration. *Personnel Psychology.* 1991; 44(3): 473-86.
16. Cousineau D, Chartier S. Outliers detection and treatment: a review. *International Journal of Psychological Research.* 2010; 3(1): 58-67.
17. Osborne JW. *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data.* Sage, 2012.
18. Osborne JW. Improving your data transformations: Applying the Box-Cox transformation. *Practical Assessment, Research & Evaluation.* 2010; 15(12): 1-9.
19. Tukey JW. On the comparative anatomy of transformations. *The Annals of Mathematical Statistics.* 1957: 602-32.
20. Osborne J. Notes on the use of data transformations. *Practical Assessment, Research & Evaluation.* 2002; 8(6): 1-8.
21. Watson S, Gallagher P, Del-Estal D, Hearn A, Ferrier IN, Young AH. Hypothalamic-pituitary-adrenal axis function in patients with chronic depression. *Psychol Med.* 2002; 32(6): 1021-8.
22. Watson S, Gallagher P, Ferrier IN, Young AH. Post-dexamethasone arginine vasopressin levels in patients with severe mood disorders. *J Psychiatr Res.* 2006; 40(4): 353-9.
23. Watson S, Gallagher P, Ritchie JC, Ferrier IN, Young AH. Hypothalamic-pituitary-adrenal axis function in patients with bipolar disorder. *Br J Psychiatry.* 2004; 184: 496-502.
24. Nagelkerke NJD. A note on a general definition of the coefficient of determination. *Biometrika.* 1991; 78(3): 691-2.
25. Pruessner JC, Wolf OT, Hellhammer DH, Buske-Kirschbaum A, von Auer K, Jobst S, et al. Free cortisol levels after awakening: a reliable biological marker for the assessment of adrenocortical activity. *Life Sci.* 1997; 61(26): 2539-49.

26. Pruessner JC, Kirschbaum C, Meinlschmid G, Hellhammer DH. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. *Psychoneuroendocrinology*. 2003; 28(7): 916-31.
27. Osborne JW. Normality of residuals is a continuous variable, and does seem to influence the trustworthiness of confidence intervals: A response to, and appreciation of, Williams, Grajales, and Kurkiewicz (2013). *Practical Assessment, Research, and Evaluation*. 2013; 18(12): 1-9.
28. Razali NM, Wah YB. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of Statistical Modeling and Analytics*. 2011; 2(1): 21-33.
29. Schmider E, Ziegler M, Danay E, Beyer L, Bühner M. Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*. 2010; 6(4): 147.
30. Miller R, Plessow F. Transformation techniques for cross-sectional and longitudinal endocrine data: Application to salivary cortisol concentrations. *Psychoneuroendocrinology*. 2013; 38(6): 941-6.
31. Hwang Gu S-L, Gau SS-F, Tzang S-W, Hsu W-Y. The ex-Gaussian distribution of reaction times in adolescents with attention-deficit/hyperactivity disorder. *Research in developmental disabilities*. 2013; 34(11): 3709-19.
32. Rigby RA, Stasinopoulos DM. Smooth centile curves for skew and kurtotic data modelled using the Box–Cox power exponential distribution. *Statistics in medicine*. 2004; 23(19): 3053-76.
33. Heathcote A, Popiel SJ, Mewhort DJ. Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*. 1991; 109(2): 340.
34. Preacher KJ, Curran PJ, Bauer DJ. Computational tools for probing interactions in multiple linear regression, multilevel modeling, and latent curve analysis. *Journal of Educational and Behavioral Statistics*. 2006; 31(4): 437-48.
35. Tukey JW. One Degree of Freedom for Non-Additivity. *Biometrics*. 1949; 5(3): 232-42.
36. Freedman DA. Bootstrapping regression models. *The Annals of Statistics*. 1981; 9(6): 1218-28.
37. McCullagh P. Generalized linear models. *European Journal of Operational Research*. 1984; 16(3): 285-92.
38. Kruschke JK. Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*. 2013; 142(2): 573.

## **Acknowledgment**

This research was financially supported by the Northumberland Tyne and Wear NHS Trust who were not involved in the collection, analysis and interpretation of data; in the writing of the report; nor in the decision to submit the article for publication